

PENENTUAN PENULARAN COVID-19 BERDASARKAN GEJALA KANDIDAT COVID-19 MEMANFAATKAN ALGORITMA NAÏVE BAYES

Yeffry Handoko Putra^{1*}, Rahma Wahdiniwati¹, Noorihan Abdul Rahman²
dan Zuriani Ahmad Zukarnain²,

¹Universitas Komputer Indonesia,

²Universiti Teknologi MARA, Cawangan Kelantan, Malaysia

Email : *yeffryhandoko@email.unikom.ac.id, rahma@email.unikom.ac.id,

noorihan@uitm.edu.my, zurianiaz@uitm.edu.my

Abstrak

Indonesia mengalami penularan covid-19 sejak awal tahun 2019 dan untuk mengatasinya pemerintah Indonesia memilih cara mitigasi dengan memberlakukan sistem pembatasan sosial (PSBB) dan kemudian diganti dengan pembaasan aktivitas (PPKM darurat dan PPKM Level 4). Mitigasi dilakukan karena sampai saat ini belum jelas gejala pasti yang menandai seseorang tertular Covid-19. Namun pada penelitian ini ingin dicoba untuk mengatasi dari sisi pencegahan yaitu dengan menentukan seberapa besar gejala-gejala dari pasien yang bisa menjadi gejala kandidat untuk Covid-19. Penentuan gejala kandidat ini dilakukan dengan memilih gejala dominan dan prioritas gejala melalui survei di beberapa puskesmas di Kabupaten Bandung Barat. Sedangkan relasi pasangan gejala kandidat terhadap penularan Covid-19 dilakukan mempergunakan algoritma Naïve Bayes. Hasil dari perhitungan data uji untuk model usulan ini diperoleh akurasi sebesar 85% dan AUC sebesar 0,878 yang menunjukkan model usulan untuk penularan Covid-19 dengan empat gejala kandidat sudah memenuhi klasifikasi baik. Model usulan ekperimental ini bisa menjadi alternatif sebelum pihak pemerintah yang berwenang dengan kesehatan belum mengumumkan kepastian gejala-gejala apa saja yang menjadi gejala pasti yang menandai seseorang tertular Covid-19.

Kata Kunci : Covid-19, Naïve Bayes, Gejala, Kandidat

1. PENDAHULUAN

Kasus penyebaran Covid-19 sudah dimulai di benua Asia sejak tahun 2019, di Indonesia sendiri Covid-19 masuk pada Februari 2019. Pemerintah Indonesia telah melakukan usaha mitigasi dengan memberlakukan pembatasan sosial (PSBB), hingga pembatasan aktivitas (PPKM Darurat dan PPKM Level 4). Pada dasarnya usaha mitigasi ini merupakan anjuran dari WHO yang didasari oleh jumlah orang yang terinfeksi, dirawat di rumah sakit dan meninggal karena Covid-19 (Instruksi Menteri Dalam Negeri Tentang Pemberlakuan Pembatasan Kegiatan Masyarakat). Sepintas keputusan ini adalah keputusan yang ada untuk kasus mengurangi penyebaran, tetapi itu tidak selalu cara yang benar. Karena pencegahan lebih penting daripada mitigasi. Pencegahan dapat dilakukan ketika kita mengetahui gejala Covid-19, namun masalah utamanya adalah tidak ada acuan yang dapat menentukan secara pasti gejala penyebab Covid-19 yang sebenarnya. Setidaknya dalam penelitian ini gejala dominan (Hikmawati and Setiyabudi) penderita COVID-19 diperoleh melalui survei kepada pemangku kepentingan terkait. Namun masalahnya gejala COVID-19 ditemukan juga pada penyakit flu biasa sehingga sulit untuk menarik garis tegas untuk gejala pasti penyebab Covid-19. Untuk alasan pencegahan, pemerintah perlu memilih dan sekaligus menentukan gejala kandidat penderita Covid-19 agar pencegahan bisa dilakukan.

Beberapa penelitian telah menggunakan algoritma K-mean untuk mengklasifikasikan data pasien COVID-19 di Indonesia, namun penelitian ini masih berfokus pada data mitigasi, bukan pada pencegahan. (Abdullah, Susilo and Saleh Ahmar). Dwitri et.all (Dwitri, Tampubolon and Prayoga) menggunakan k-means untuk mengelompokkan data COVID-19 untuk mengetahui penyebaran penyakit di Indonesia. Penyebaran centroid hasil clustering dengan k-means meliputi wilayah Jakarta, Jawa Barat, dan Banten. Jakarta salah satu daerah yang memiliki penyebaran data COVID-19 yang cukup tinggi, Jawa Barat, potensial sedangkan Banten cukup potensial. Penelitian yang dilakukan oleh Untoro dkk

(Cahyo Untoro, Anggraini and Andini) menyimpulkan bahwa jumlah kluster yang diperoleh dari data penyebaran virus COVID-19 di Jakarta hanya dua kluster, padahal pemerintah Indonesia telah menerapkan 4 level aturan pembatasan aktivitas (PPKM).

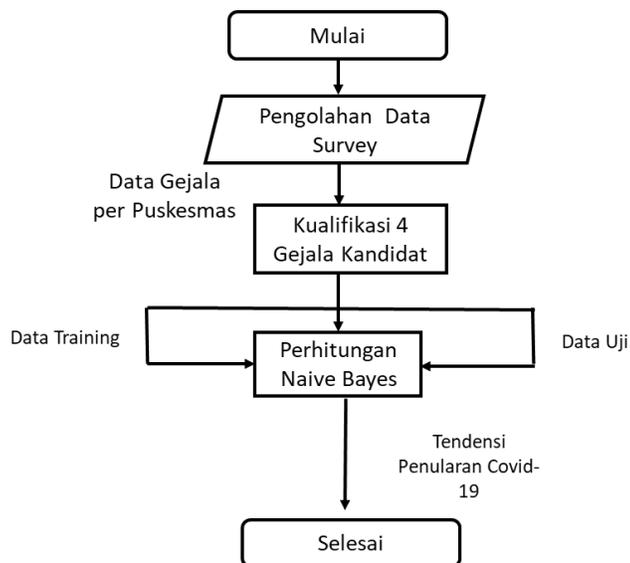
Hingga saat ini, belum ada yang meneliti persentase maksimum kemungkinan jumlah orang yang memiliki gejala mirip COVID-19 justru akan menderita COVID-19. Namun setidaknya jumlah penderita dengan gejala mirip COVID-19 ini bisa dimanfaatkan oleh pemerintah atau pihak terkait untuk mulai melakukan tindakan preventif dan menghindari lockdown. Seperti yang kita ketahui, adanya lock down atau pembatasan bersyarat menyebabkan perekonomian masyarakat menurun (Taruna Habir and Wardana), terutama bagi masyarakat yang bekerja di sektor pabrik, berwiraswasta tanpa gaji bulanan yang tetap. Beberapa tulisan telah membuktikan bahwa pandemi COVID-19 telah menyerang perekonomian masyarakat di Indonesia dan telah menurunkan daya beli dan pendapatan masyarakat. Ditambah lagi dengan penerapan lockdown di mall, sekolah, perkantoran, pabrik dan pusat hiburan (Update on PPKM Policy in Indonesia Effective as of March 22, 2022), yang membatasi ruang gerak masyarakat dalam meningkatkan pendapatannya.

Penelitian ini ingin mencoba melakukan (1)Pemilihan gejala kandidat Covid-19 (2) Penentuan prioritas gejala kandidat Covid-19 (3) Penentuan pasangan gejala kandidat yang memiliki kecendrungan terhadap penularan Covid-19.

2. METODOLOGI

2.1. Tahapan Penelitian

Untuk mengetahui gejala yang dominan dilakukan dengan mengadakan survey di beberapa tempat di Kabupaten Bandung Barat. Setelah data survey terkumpul maka didapatkan kandidat gejala penyebab Covid-19. Kemudian kandidat gejala hasil survei di 11 Puskesmas ini dibagi menjadi data latih dan data uji untuk algoritma Naïve Bayes dalam menentukan penularan Covid-19 seperti terlihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Dari hasil survey awal diperoleh 6 gejala terbesar yang diduga diderita oleh penderita Covid-19, kemudian dari 6 gejala tersebut ditentukan prioritas gejala pada penderita Covid-19 berdasarkan survey lanjutan. Keenam gejala tersebut diperlihatkan pada tabel 1. Kemudian secara kualitatif, masing-masing gejala ini tingkatannya dibagi secara rasional berdasarkan keluhan dari pasien. Yang kemudian data ini dijadikan data latih dan data uji untuk algoritma Naïve Bayes.

Tabel 1. Kandidata Gejala Penderita Covid-19 dengan Prioritasnya

No. Gejala	Nama Gejala	Prioritas
1	Demam Tinggi	3
2	Batuk Kering	2
3	Pundak nyeri	6
4	Sakit Tenggorokan	1
5	Sakit Kepala	5
6	Sesak Napas	4

2.2. Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa lalu sehingga dikenal sebagai algoritma Naïve Bayes.

Menurut Liu (Mujiexin , Hongjun and Tian Rui) Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

Keterangan :

- X = Data dengan class yang belum diketahui
- H = Hipotesis data X merupakan suatu class spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (porteriori prob.)
- P(H) = Probabilitas hipotesis H (prior prob.)
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X

Untuk menjelaskan teorema Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, Menurut Saleh [11] teorema bayes di atas disesuaikan sebagai berikut :

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(C)} \tag{2}$$

Adapun alur dari algoritma Naive Bayes adalah sebagai berikut :

1. Baca data training
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:
 - a. Cari nilai mean dan standar deviasi dari masing masing parameter yang merupakan data numerik.
 - b. Cari nilai probabilitistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel mean, standart deviasi dan probabilitas

Dimana variabel C merepresentasikan kelas, sementara variabel F1 ... Fn merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C.

2.3. Algoritma Evaluasi dan Validasi

Validasi adalah proses mengevaluasi keakuratan hasil model prediksi. K-Fold Cross Validation merupakan teknik validasi yang membagi data menjadi k bagian dan kemudian setiap bagian akan dilakukan proses klasifikasi. Dengan menggunakan K-Fold Cross Validation akan dilakukan percobaan sebanyak k. Setiap eksperimen akan menggunakan satu data testing dan bagian K-1 akan menjadi data training, kemudian data testing tersebut akan ditukar dengan salah satu data training, sehingga untuk setiap eksperimen akan mendapatkan data testing yang berbeda. Untuk menguji model, digunakan metode Matriks Konfusi, dan kurva ROC.

1. Matriks Konfusi digunakan untuk menganalisis seberapa baik classifier dapat mengidentifikasi tupel dari kelas yang berbeda.

2. Kurva ROC,

Kurva ROC (Receiver Operating Characteristic) menunjukkan akurasi klasifikasi dan membandingkan secara visual. ROC mengungkapkan matriks kebingungan. ROC adalah grafik dua dimensi dengan positif palsu sebagai garis horizontal dan positif benar untuk mengukur perbedaan kinerja metode yang digunakan. Kurva ROC digunakan untuk mengukur AUC (Area Under the Curve). AUC dihitung untuk mengukur perbedaan kinerja metode yang digunakan (Gorunescu). Kurva ROC membagi hasil positif pada sumbu y dan hasil negatif pada sumbu x ((Gorunescu), (Witten, Ian H; Frank, Eibe; Hall, Mark A; Pal, Christopher J)). Jadi semakin besar area di bawah kurva, semakin baik hasil prediksinya.

Nilai AUC dibagi menjadi beberapa kelompok:

0,90 -1,00 = Klasifikasi Sangat Baik

0,80 -0,90 = Klasifikasi Baik

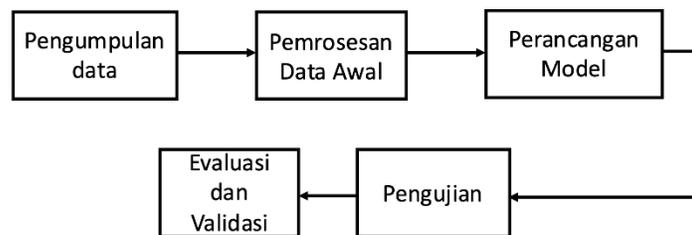
0,70 -0,80 = Klasifikasi Sedang

0,60 -0,70 = Klasifikasi Buruk

0,50 -0,60 = Gagal

2.4. Tahapan Penelitian

Metode yang yang digunakan adalah metode penelitian eksperimental dengan tahapan penelitian diperlihatkan pada gambar 2.



Gambar 2. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Dari hasil survey kasar (sekitar 6 bulan selama tahun 2020) di 11 Kabupaten Bandung Barat di Puskesmasnya seperti terlihat pada Tabel 2, ditemukan 6 gejala dominan yang mirip dengan gejala penderita Covid-19, yang kemudian disebut sebagai kandidat gejala Covid-19 yaitu: demam tinggi, batuk kering, nyeri bahu, sakit tenggorokan, sakit kepala, dan sesak napas. Dari gejala tersebut ditentukan level dari keparahan suatu gejala, meskipun gejala yang ringan belum tentu tidak ditemui pada penderita Covid-19. Dari hasil survey, urutan dominasi mulai dari yang tertinggi hingga terendah untuk gejala kandidat penderita Covid-19 adalah sebagai berikut: sakit tenggorokan, batuk kering, demam tinggi, sesak napas, sakit kepala, dan bahu. rasa sakit.

Tentunya urutan dominasi gejala yang mirip dengan gejala penderita Covid-19 hanya lewat di lokasi penelitian yang diamati, sedangkan untuk daerah lain perlu dilakukan survei lagi. Urutan dominasi gejala tersebut nantinya akan digunakan untuk menentukan 4 gejala kandidat penderita Covid-19 (Tabel 3).

Tabel 2. Jumlah gejala yang ditemui per 50 orang penderita

No	Puskesmas	Gejala 1	Gejala 2	Gejala 3	Gejala 4	Gejala 5	Gejala 6
1	Jaya Mekar	4	3	2	4	1	3
2	Tagogapu	1	2	3	5	5	5
3	Padalarang	2	4	2	1	3	4
4	Nglamprah	4	1	5	4	3	4
5	Cimareme	2	1	5	2	3	4
6	Padasuka	1	2	1	1	3	1
7	Rajamanda	2	2	1	1	1	2
8	Cipendeuy	5	2	5	2	1	3
9	Cipatat	5	3	5	2	4	3
10	Batujajar	4	5	1	3	1	2
11	Cihampelas	3	5	4	3	2	1

Pemerintah Indonesia menetapkan aturan lock down berdasarkan tingkat jumlah orang yang terkena dampak pandemi COVID-19 pada level terberat, misal pada PPKM adalah level 3 dan 4, namun sekali lagi ini merupakan penanganan mitigasi. Jika dilihat dari segi pencegahan maka dilakukan investigasi yang lebih mendalam yaitu gejala apa saja yang ada di masyarakat yang harus dijadikan perhatian penting sebelum melakukan *lock down*. Pendekatan pencegahan ini lebih awal yaitu dengan memperhatikan seberapa besar kemungkinan gejala dominan mirip penderita Covid-19 pada seseorang yang dimiliki oleh penderita Covid-19 atau akan menyebabkan seseroang menjadi penderita Covid-19 di seminggu kemudian. Lalu dipilih 4 gejala dominan sebagai kandidat gejala penderita Covid-19 (Tabel 3).

Tabel 3. Empat Kandidat Gejala Penderita Covid-19 Serta Kesesuaiannya Dengan Gejala Pada Penderita Covid-19

Kasus	Gejala 1	Gejala 2	Gejala 3	Gejala 4	Ditemui pada penderita Covid-19
1	Tidak sering	Ringan	Ringan	Sedang	Tidak
2	Tidak sering	Ringan	Ringan	Menengah	Tidak
3	Tidak sering	Ringan	Berat	Sedang	Tidak
4	Tidak sering	Berat	Ringan	Sedang	Tidak
5	Tidak sering	Berat	Berat	Sedang	Ya
6	Tidak sering	Berat sekali	Ringan	Sedang	Ya
7	Tidak sering	Berat sekali	Berat	Sedang	Ya
8	Tidak sering	Berat sekali	Berat	Menengah	Ya
9	Sering	Ringan	Ringan	Sedang	Ya
10	Sering	Ringan	Ringan	Menengah	Ya
11	Sering	Ringan	Berat	Menengah	Ya
12	Sering	Berat	Ringan	Sedang	Ya
13	Sering	Berat	Ringan	Menengah	Ya
14	Sering	Berat	Berat	Sedang	Ya
15	Sering	Berat sekali	Ringan	Sedang	Ya
16	Sering	Berat sekali	Berat	Menengah	Ya

Tabel 4. Frekuensi gejala ditemui pada penderita Covid-19

	Gejala pada	
	Jumlah	Peluang
Ya	12	0,75
Tidak	4	0,25

Proses pengujian perlu dilakukan koreksi saat peluangnya nol yaitu pempergunakan koreksi Laplacian seperti pada contoh di tabel 5.

Tabel 5. Contoh Pengujian

Contoh Pengujian	Gejala 1 Sering	Gejala 2 Berat sekali	Gejala 3 Ringan	Gejala 4 Sedang
Tanpa Koreksi				
Peluang Ya	2,000	0,417	0,500	0,583
Peluang Tidak	0,000	0,000	0,750	0,750
Dengan Koreksi (<i>Laplacian</i>)				
Peluang Ya	2,000	0,417	0,500	0,583
Peluang Tidak	0,250	0,250	0,750	0,875
Keputusan				
Total Peluang Ya	0,243	1		
Total Peluang Tidak	0,041	2		

Hasil pengujian model akan dibahas melalui matriks konfusi (Tabel 6) untuk menunjukkan seberapa baik model yang terbentuk.

Tabel 6. Matriks Konfusi Model Naïve Bayes Penularan Covid

	True Positif	True Negatif	Presisi
Prediksi Positif	90	21	81%
Prediksi Negatif	10	79	89%
Class Recall	90%	79%	

Menggunakan Tabel 6 dan menghitung nilai akurasi dan AUC (Tabel 7), dapat diketahui bahwa pengujian menggunakan model klasifikasi Naïve Bayes pada penelitian ini menghasilkan klasifikasi yang baik, dengan nilai akurasi sebesar 85%, dan nilai AUC sebesar 0,878 yang termasuk Klasifikasi Baik

Tabel 7. Validasi Model Naïve Bayes

Klasifikasi Tinjauan Negatif yang Berhasil	Klasifikasi positif yang berhasil	Akurasi Model	AUC
90%	79%	85%	0,878

4. KESIMPULAN

1. Dari data yang diberikan diperoleh metoda untuk menentukan kandidat gejala Covid-19 yang diperoleh untuk daerah sampling Kabupaten Bandugn Barat memanfaatkan algoritma Naïve Bayes diperoleh Model penentuan penularan Covid-19 bedasarkan empat gejala kandidat memiliki presisi 89% dan akurasi 85%. Meskipun akurasi dan presisi belum mencapai nilai 90%

tetapi berdasarkan klasifikasi dari (Gorunescu) sudah termasuk Klasifikasi Baik dengan nilai AUC 0,878

2. Penentuan penularan Covid-19 merupakan model usulan alternatif disaat pihak pemerintah yang berwenang dengan kesehatan belum mengumumkan kepastian gejala-gejala apa saja yang menjadi gejala-gejala pasti yang menandai seseorang tertular Covid-19

DAFTAR PUSTAKA

- Taruna Habir, Manggi and Wisnu Wardana. *COVID-19's Impact on Indonesia's Economy and Financial Markets*. Singapore: ISEAS Yusof Ishak Institute, 2020. <https://www.iseas.edu.sg/wp-content/uploads/2020/12/ISEAS_Perspective_2020_142.pdf>.
- Update on PPKM Policy in Indonesia Effective as of March 22, 2022*. Jakarta: Seek Legal Consultant, 2022.
<https://www.ssek.com/images/Update_on_PPKM_Policy_in_Indonesia_Effective_as_of_March_22_2022.pdf>.
- Abdullah, Dahlan, et al. "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data." *Quality & Quantity* 56.3 (2022): 1283-1291. <<https://pubmed.ncbi.nlm.nih.gov/34103768/>>.
- Dwitri, N, et al. *Jurnal Teknologi Informasi* 4.1 (2020): 128-132.
- Cahyo Untoro, Meida, et al. "Implementation of k-means method for COVID-19 data clustering in Jakarta Province." *Jurnal Ilmiah Sistem Informasi* 11.2 (2021): 58-88.
- Hikmawati, Isna and Ragil Setiyabudi. "Epidemiology of COVID-19 in Indonesia: common source and propagated source as a cause for outbreaks." *Journal of Infection in Developing Countries* 15.5 (2021): 646-652.
- Instruksi Menteri Dalam Negeri Tentang Pemberlakuan Pembatasan Kegiatan Masyarakat*. Jakarta: Indonesian Ministry of Home Affairs, 2021.
- Mujiexin , Liu, et al. "Naive Bayes clusterer." *Conference on Data Science and Knowledge Engineering for Sensing Decision Support*. 2018.
- Witten, Ian H; Frank, Eibe; Hall, Mark A; Pal, Christopher J. *Data Mining: Practical Machine Learning Tools and Techniques, Fourth Edition*. Cambridge, USA: Morgan Kaufman, 2017.
- Han, Jiawei; Kamber, Micheline;. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2006.
- Gorunescu, Florin. *Data Mining: Concepts, Models and Techniques*. Springer-Verlag Berlin Heidelberg, 2011.